Face Recognition Evaluation @ Idemia

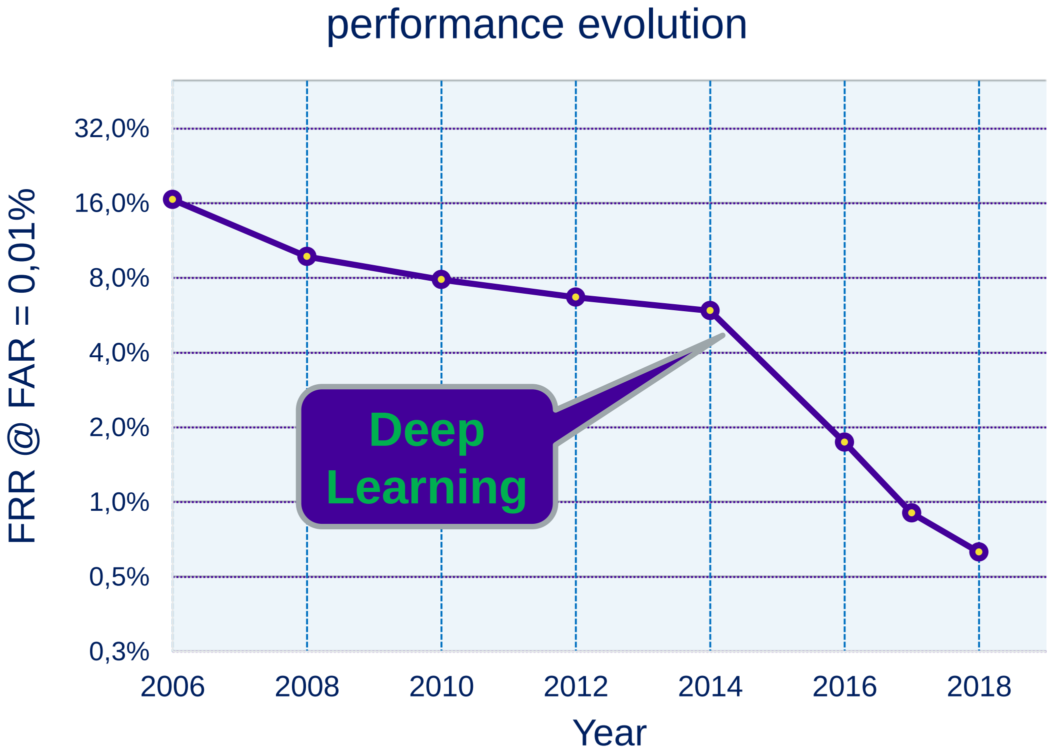
Stephane Gentric Ph.D.
Research Unit Manager
Senior Expert

IDEMIA

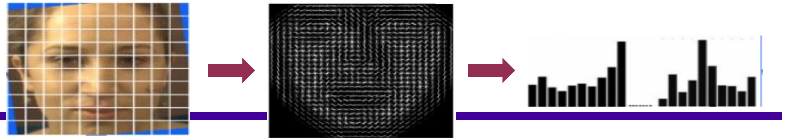
Face Recognition Performance

A long history of FR evaluation at Sagem / L1 / Safran / Morpho / Idemia

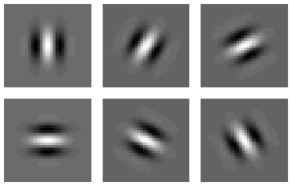
- **Median value on 40 different databases.**
- **Absolute performance varies with images quality**

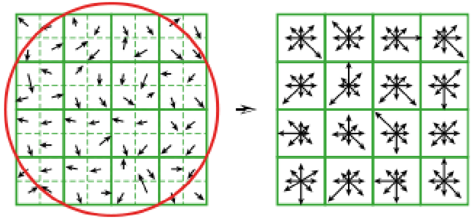


Traditional Approaches

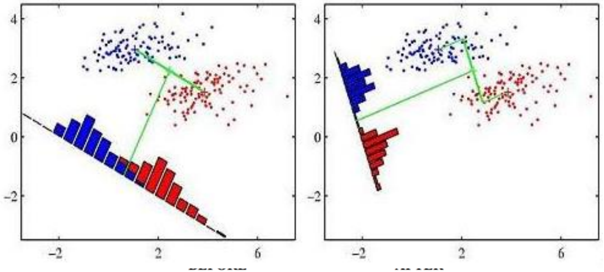


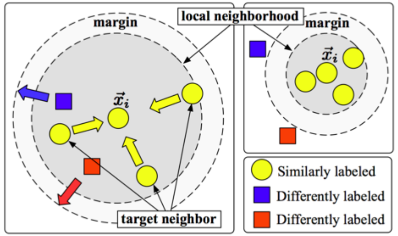
These methods are composed of two steps





1. Agnostic features are extracted from the images
   - Most of the discriminative information is born by the gradients
   - Various representations of the gradients have been proposed
     - SIFT, HOG, LBP
     - Filter bank responses (Gaussian derivatives, Gabor, etc.)
2. A transformation of the initial feature space is learned
   - It reduces the dimensionality and concentrates the discriminative information (remove noise and redundancy)
   - Most methods simply use a linear transformation
     - LDA, ITML, LMNN, …
   - But more complex methods exist (e.g. local metric)





Deep Networks for Face Recognition

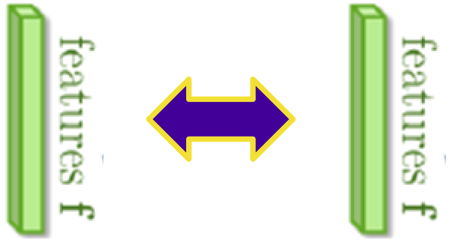
Coding = build a template from an image

A Face template is learned directly from the pixels.

- **A chain of forward operation in the Network**
- **Coding is done on CPU or GPU. On a PC, a Smartphone or in the Cloud**

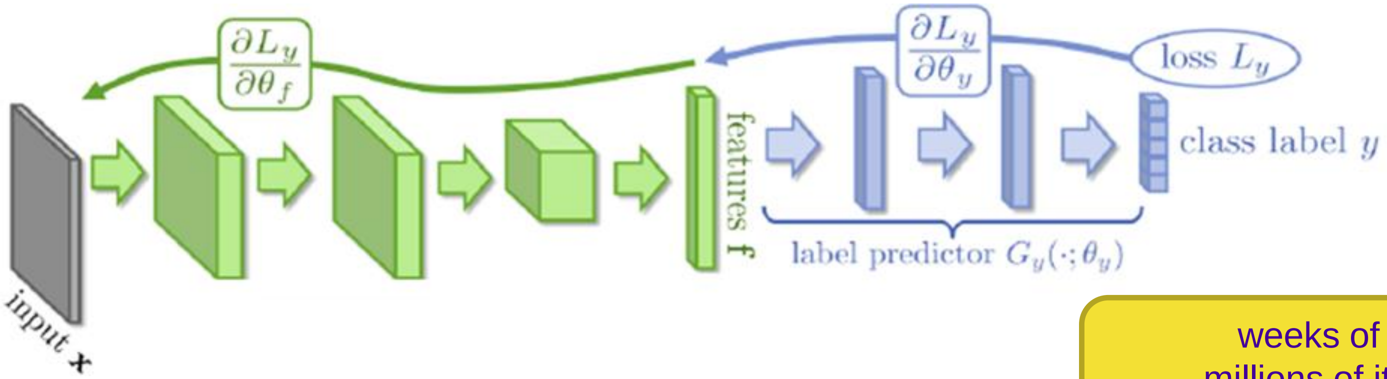
Matching = compare 2 templates

- **A simple Metric is used**
- **Very fast. Millions per second per core**



Deep Networks for Face Recognition

- A Deep Network is learned by minimizing iteratively a loss Function.

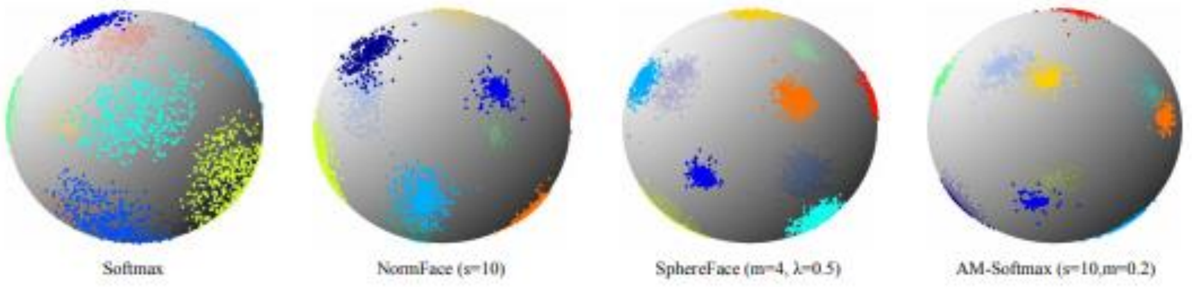


weeks of GPUs
millions of iterations
millions of images

Biometric Performances comes from :

- **The size of the image database** (number of identities, number of images per id)
- **The architecture of the Network (number of layers, maps …)**
- **The loss functions**

Finding the correct loss function



- SoftMax Cross Entropy

$$\mathcal{L}_{Softmax} = -\sum_{i=1}^{N} log \frac{exp(\mathbf{w}_{y_i}^T f_i + b_{y_i})}{\sum_{l=1}^{M} exp(\mathbf{w}_l^T f_i + b_l)}$$

- Center Loss

A discriminative feature learning approach for deep face recognition, Y.Wen & all

$$\mathcal{L}_C = \frac{1}{2} \sum_{i=1}^{m} ||x_i - c_{y_i}||_2^2$$

- Additive Margin:

Additive Margin Softmax for Face Verification, F.Wang & all

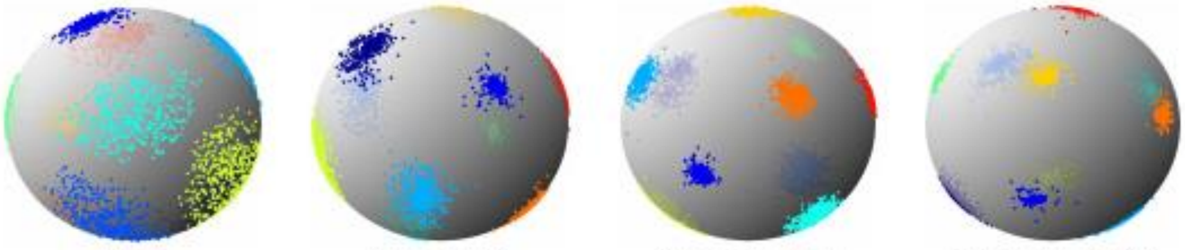
$$\mathcal{L}_{AMS} = -\frac{1}{n} \sum_{i=1}^{n} log \frac{e^{s\cdot(cos\theta_{y_i} - m)}}{e^{s\cdot(cos\theta_{y_i} - m)} + \sum_{j=1, j\neq y_i}^{c} e^{s\cdot cos\theta_j}}$$

- SphereFace

SphereFace: Deep Hypersphere Embedding for Face Recognition, W.Lui & all

$$L_{\text{ang}} = \frac{1}{N} \sum_i - \log\left(\frac{e^{\|\boldsymbol{x}_i\| \cos(m\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\| \cos(m\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\| \cos(\theta_{j,i})}}\right)$$

Finding the correct loss function



The most of the cited Loss functions try to project identities in some specific space region to maximize the classification on the learning database.

But:

- **We also want to control the local density in term of ethnicity, pose, gender or image quality …**
- **In other terms : Whatever the face image projected on the feature space, we want to have the same FAR (same probability to have other identity nearby)**

Finding the correct loss function

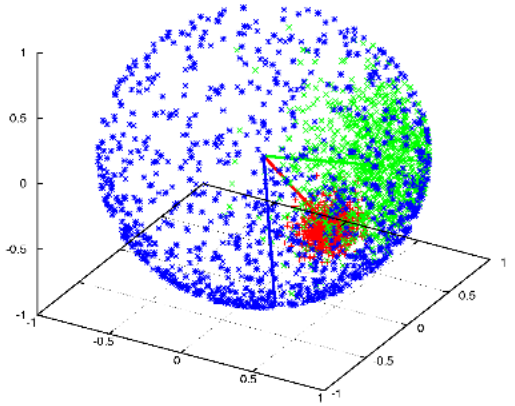
- Von Mises-Fisher Loss

Von mises-fisher mixture model-based deep learning: Application to face verification.
A.Hasnat & all @ Idemia

$$V_d(\mathbf{x}|\mu, \kappa) = C_d(\kappa)\exp(\kappa\mu^T\mathbf{x})$$

$$\mathcal{L}_{vMFML} = -\sum_{i=1}^{N} log\frac{\exp(\kappa\mu_j^T\mathbf{x}_i)}{\sum_{l=1}^{M}\exp(\kappa\mu_l^T\mathbf{x}_i)}$$

with $x = \frac{f}{\|f\|}$ and $\mu = \frac{w}{\|w\|}$



vMF distribution is a model to represent the probability of an image to belong to a person.
In the feature space, a person is represented by a center μ and a concentration κ.

This model allows a direct link between a person of the learning database and the size of the region it is supposed to take in the feature space.
It allows to control unbalanced databases in term of ethnicity, gender, age, qualities …

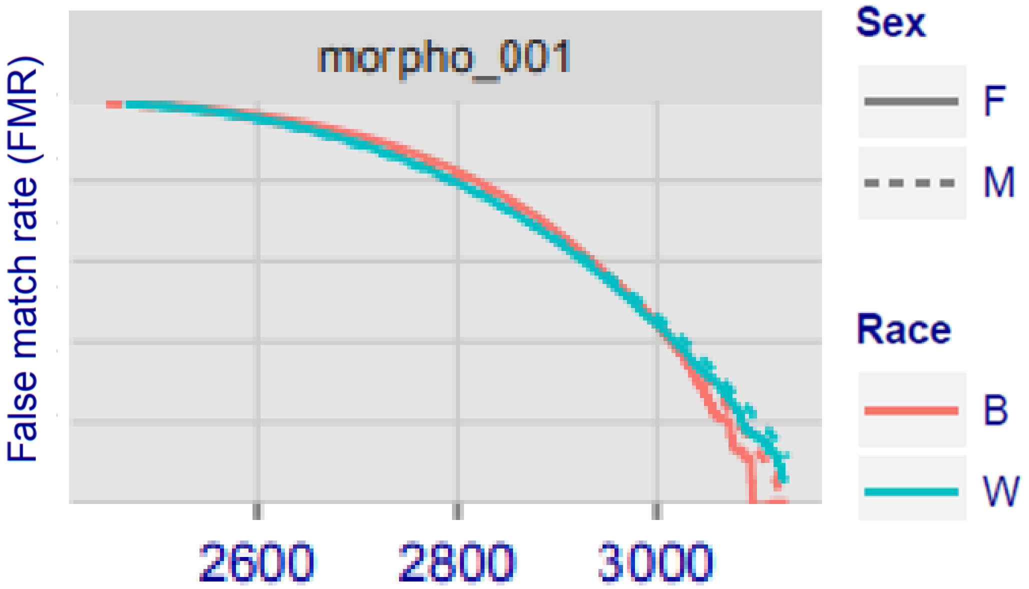
FMR Robustness
to
real live condition

- Ethnicity – Gender
- Aging
- Ambient light
- Pose
- Makeup
- Database Size

Ethnicity - Gender

- **NIST performs (FRVT 1:1 ongoing) tests regarding ethnicity and gender**

Idemia Face Recognition Algorithm have the same FMR for Black or White subjects, Male of Female.



Ethnicity - Gender

- **NIST performs (FRVT 1:N 2018) tests regarding ethnicity and gender**

Idemia Face Recognition Algorithm have the same FPIR for Black or White subjects, Male of Female, **whatever the threshold**.

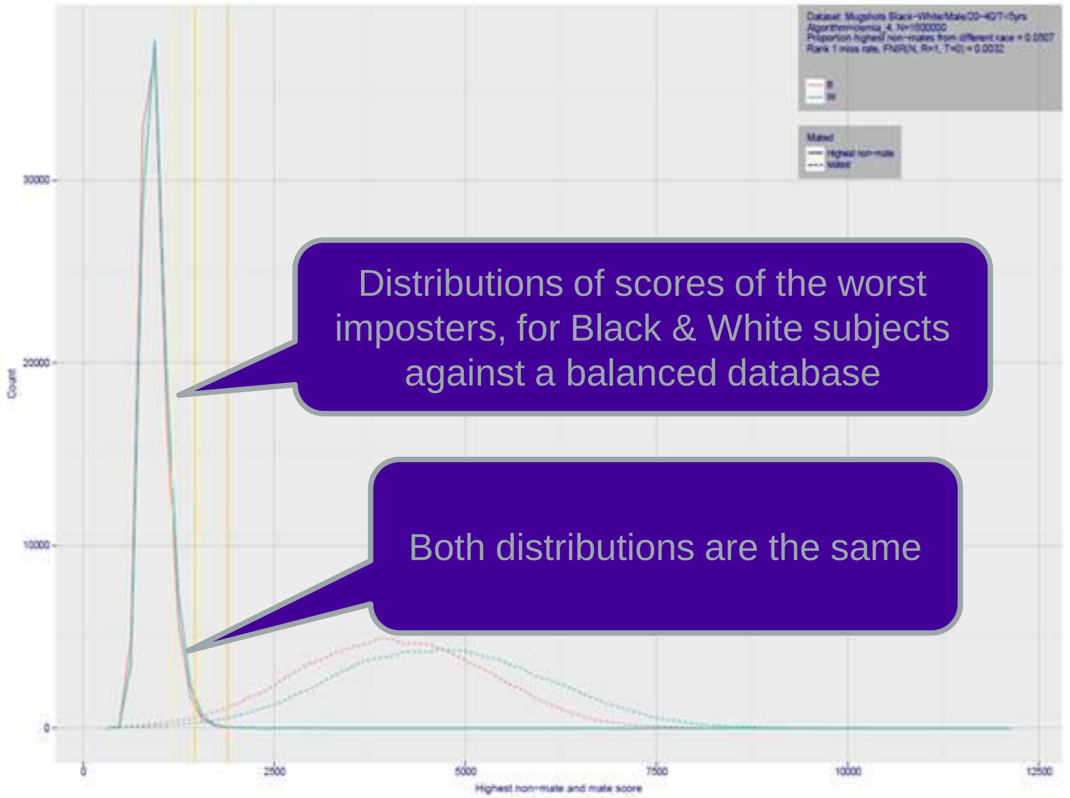


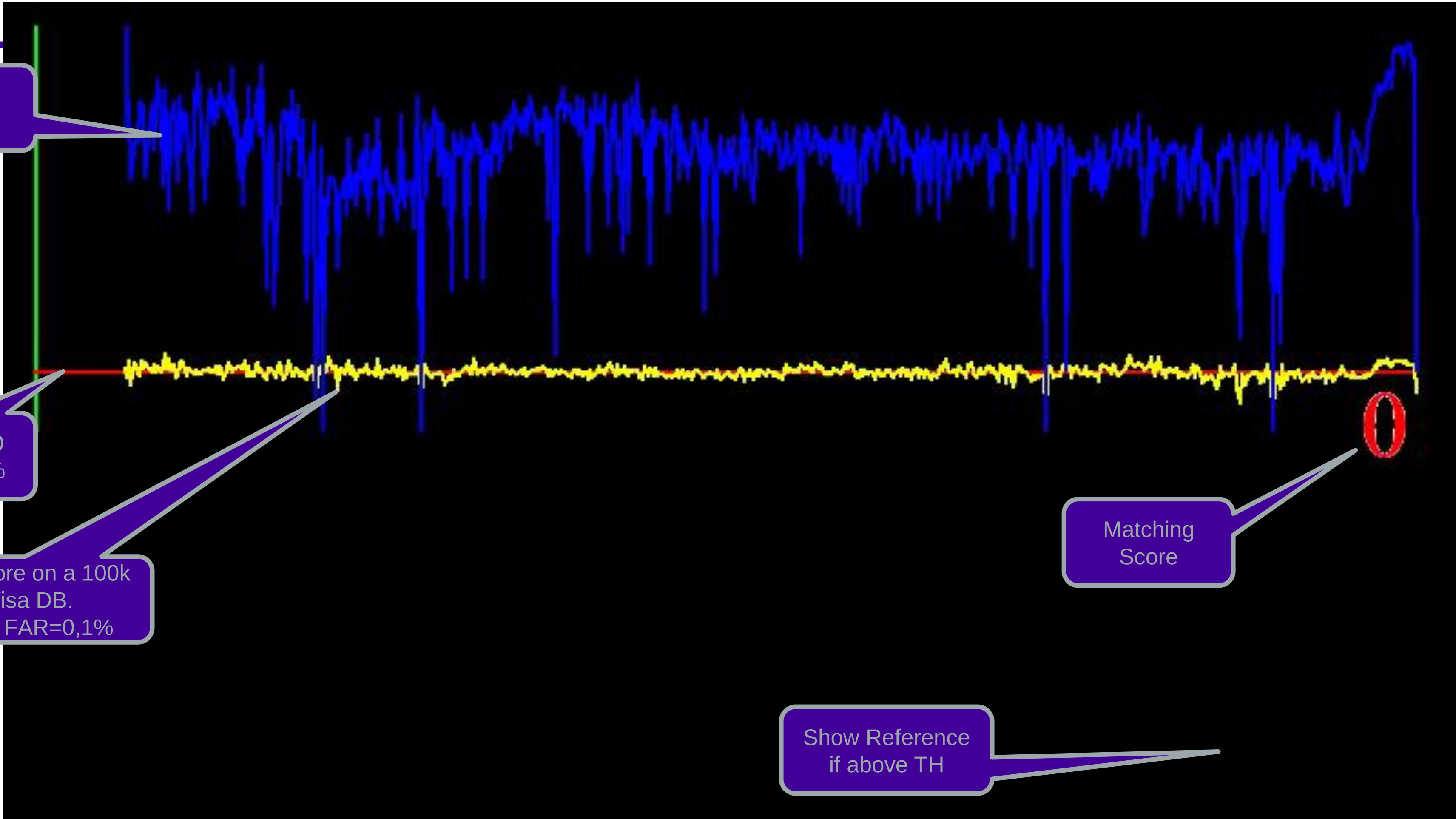
Figure 128: **[Demographic Differential: Race]** *The plots are histograms of mate scores and highest non-mate scores produced by the given algorithm when executing searches into a gallery composed of single images of equal numbers of black and white males, age 21-40. The gallery size is 1.6 million. All images are mugshots. The search image is taken in a different year to the gallery photo, and within five years, so ageing effects are constrained. The histogram bin width is the interquartile range / 20. The vertical lines correspond to thresholds that give FPIR(T) = 0.1, 0.01, 0.001. While the experimental design is intended to isolate one specific variation, other unrecorded factors may be influential; e.g. subject height.*

Aging

- Aging alters visual appearance of faces
- Aging increases probability to have modification of hair style, facial hair, wrinkles, marks …
- Aging has bigger impact on young subject

- New generation of algorithms are more robust to aging

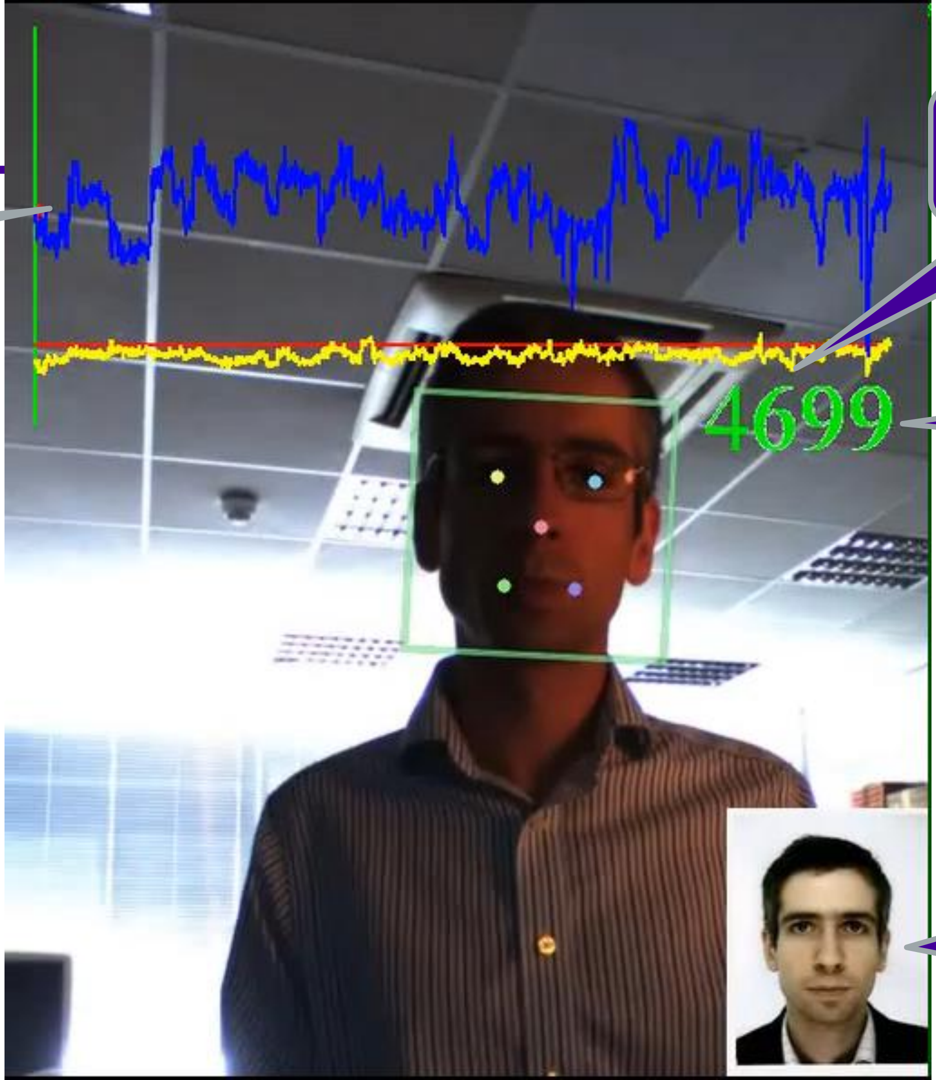
Aging : 5 years, 14 to 19



Ambient lighting

Score vs Reference

TH = 3000
FAR=0,1%

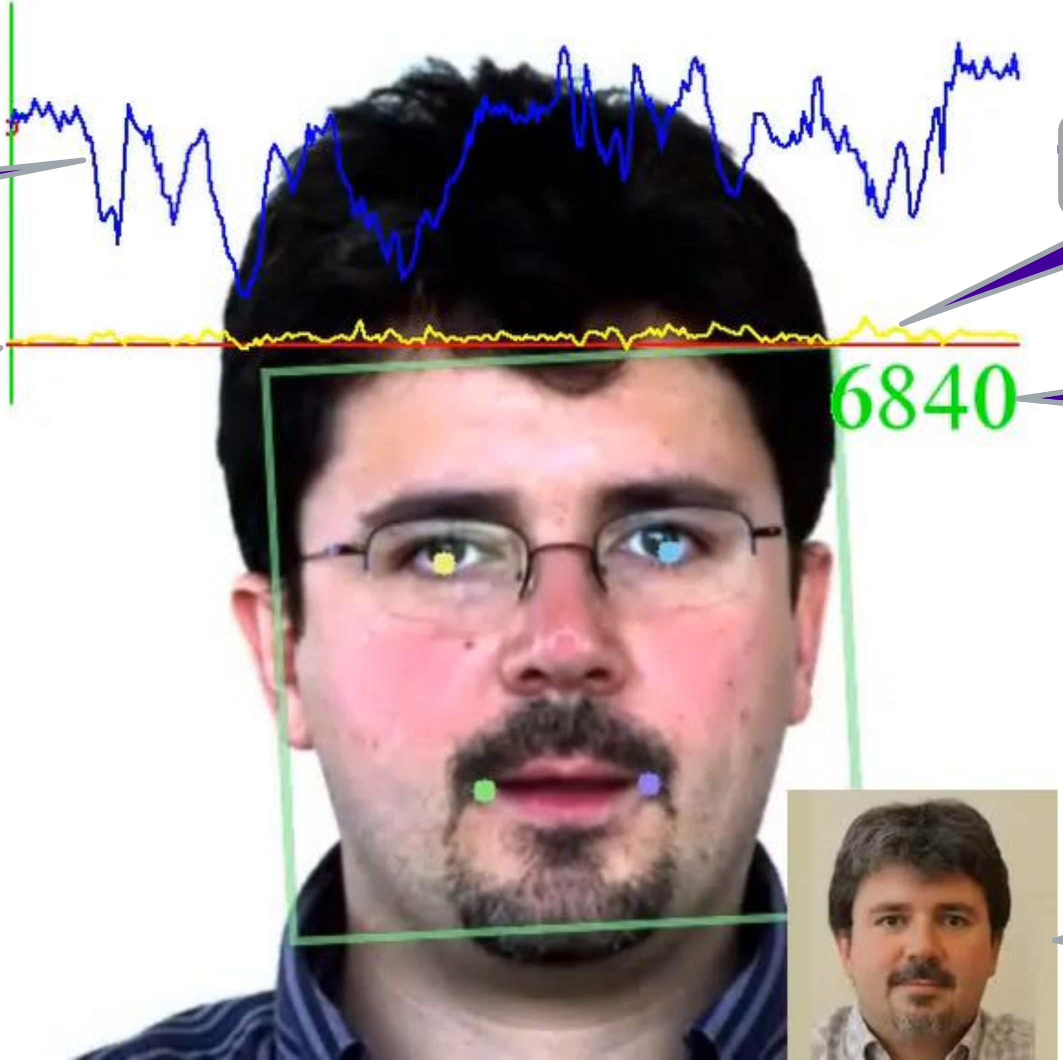


Rank 100 score on a 100k noise Visa DB. Real TH @ FAR=0,1%

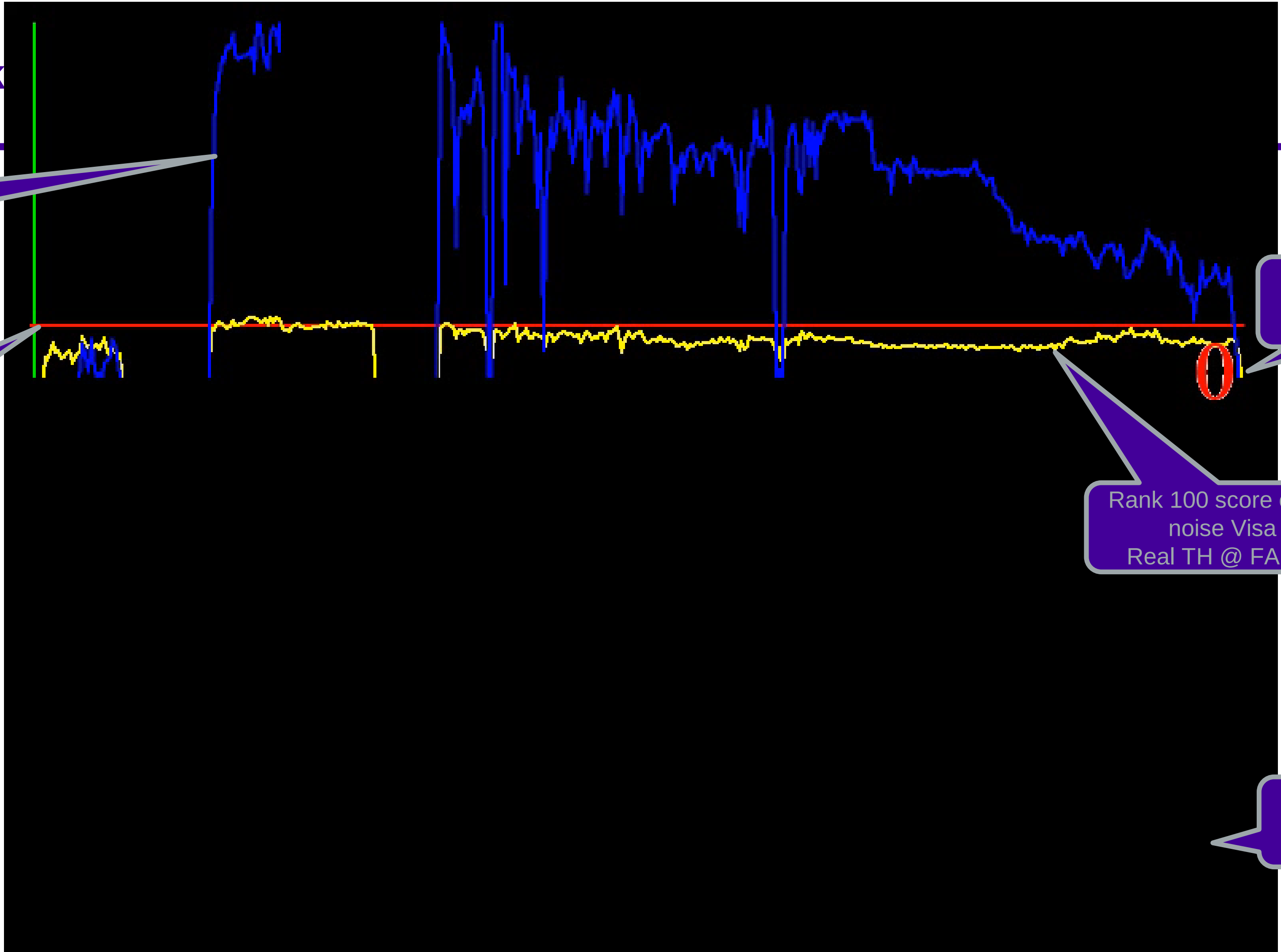
Score vs Reference

Reference if above TH

Pose


Score vs Reference
TH = 3000 FAR=0,1%
Rank 100 score on a 100k noise Visa DB. Real TH @ FAR=0,1%
6840
Score vs Reference
Reference if above TH

Mak

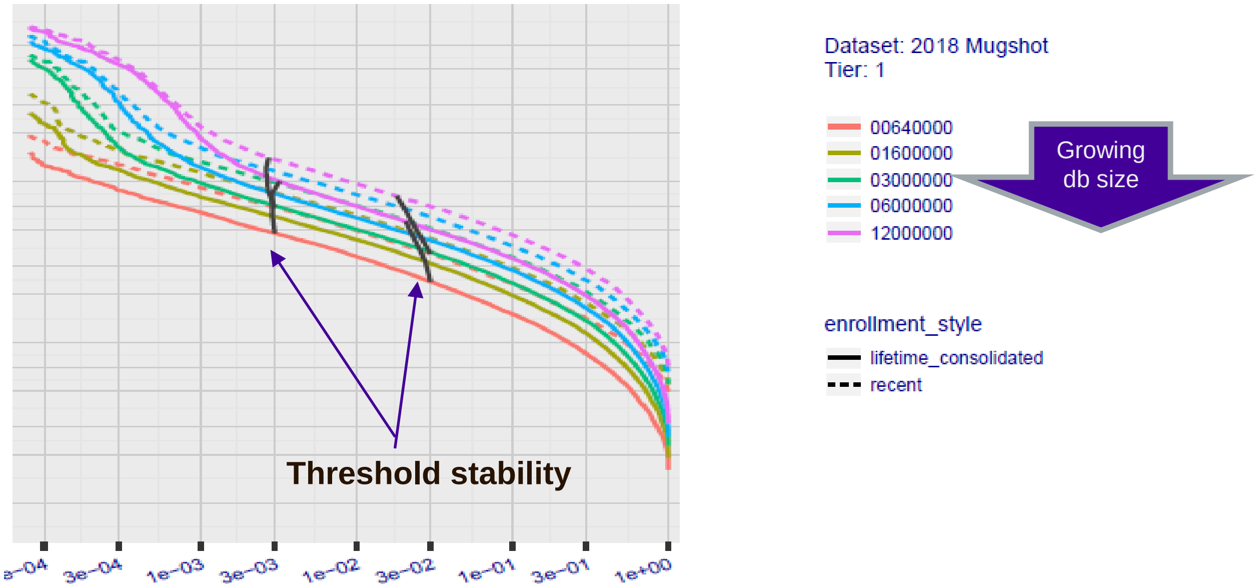


Score vs Reference
TH = 3000 FAR=0,1%
Score vs Reference
0
Rank 100 score on a 100k noise Visa DB. Real TH @ FAR=0,1%
Reference if above TH

Database Size and Threshold Stability

NIST 2018 1:N Face Recognition Benchmark

FNIR



FPIR

conclusion

- **Deep Learning is not a magic black box.**

  We just optimize parameters to perform a task on a large database.

- **Face recognition algorithms improve fast**

  By using appropriate loss functions, we improve not only the overall accuracy, but also the FMR and FPIR stability in relation to various factors.

Thank You

stephane.gentric@idemia.com